

HP StorageWorks

Clustered File System for Linux: NFS Configuration 3.0 installation and administration guide



395342-001

Part number: 395342-001
First edition: May 2005



Legal and notice information

© Copyright 1999-2005 PolyServe, Inc.

Portions © 2005 Hewlett-Packard Development Company, L.P.

Neither PolyServe, Inc. nor Hewlett-Packard Company makes any warranty of any kind with regard to this material, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. Neither PolyServe nor Hewlett-Packard shall be liable for errors contained herein or for incidental or consequential damages in connection with the furnishing, performance, or use of this material.

This document contains proprietary information, which is protected by copyright. No part of this document may be photocopied, reproduced, or translated into another language without the prior written consent of Hewlett-Packard. The information is provided "as is" without warranty of any kind and is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Neither PolyServe nor HP shall be liable for technical or editorial errors or omissions contained herein.

The software this document describes is PolyServe confidential and proprietary.

PolyServe and the PolyServe logo are trademarks of PolyServe, Inc.

PolyServe Matrix Server contains software covered by the following copyrights and subject to the licenses included in the file `thirdpartylicense.pdf`, which is included in the PolyServe Matrix Server distribution.

Copyright © 1999-2004, The Apache Software Foundation.

Copyright © 1992, 1993 Simmule Turner and Rich Salz. All rights reserved.

Copyright © 2000, Emulex Corporation. All rights reserved.

Copyright © 1994-2004, Sun Microsystems, Inc. All rights reserved.

Copyright © 1999, 2000, 2001 Lev Walkin <vlm@spelio.net.ru>. All rights reserved.

Copyright © 1996, 1997, 1998, 1999, 2000, 2001 Institut National de Recherche en Informatique et en Automatique (INRIA). All rights reserved.

Copyright © 2001 QLogic Corp. All rights reserved.

Copyright © 1993-2001 Spread Concepts LLC. All rights reserved.

Copyright © 2001 Daniel Barbalace. All rights reserved.

Copyright © 2003 Storage Networking Industry Association. All rights reserved.

Copyright © 1995-2002 Jean-loup Gailly and Mark Adler.

Linux is a U.S. registered trademark of Linus Torvalds

Microsoft, Windows, Windows NT, and Windows XP are U.S. registered trademarks of Microsoft Corporation.

Oracle® is a registered U.S. trademark of Oracle Corporation, Redwood City, California.

UNIX® is a registered trademark of The Open Group.

Clustered File System for Linux: NFS Configuration Installation and Administration Guide

Contents

HP Technical Support

HP Storage Web Site	v
HP NAS Services Web Site	v

1 Introduction

CFS-Linux Features	1
CFS-Linux Concepts and Definitions	3
Supported NFS Version	5
CFS-Linux Tested Configuration Limits	5

2 Configure Export Groups

Overview	6
Export Records and Groups	6
Virtual NFS Services	6
One or Many Export Groups?	7
High Availability and Failover Support	7
Caveats Regarding Export Groups	7
Mount PSFS Filesystems	8
Add an Export Group	9
NFS Exports Tab	10
Virtual NFS Services Tab	13
Export Group High-Availability Monitor	15
View the High-Availability Monitor	15
How the High-Availability Monitor Works	16
Advanced Options for Export Groups	17
Probe Configuration	17
Probe Severity	18
Scripts	20
View Export Group Properties	23
Modify an Export Group	24

Manage CFS-Linux from the Applications Tab	25
Other Export Group Procedures	27
Disable an Export Group	27
Enable an Export Group	27
Clear an Error Associated with an Export Group	27
Delete an Export Group	27
3 Configure Virtual NFS Services	
Overview	28
Sample Configurations	28
Guidelines for Creating Virtual NFS Services	29
Add a Virtual NFS Service	30
Migrate a Virtual NFS Service	32
From the Servers or Virtual Hosts Tab	32
From the Application Tab	33
From the Command Line	34
Other Virtual NFS Service Procedures	34
Disable a Virtual NFS Service	34
Enable a Virtual NFS Service	34
Delete a Virtual NFS Service	35
4 Other Configuration Information	
NFS Clients	36
Timeout Configuration	36
Client Mounts	36
Using the NLM Protocol	37
Index	

HP Technical Support

Telephone numbers for worldwide technical support are listed on the following HP web site: <http://www.hp.com/support>. From this web site, select the country of origin. For example, the North American technical support number is 800-633-3600.

NOTE: For continuous quality improvement, calls may be recorded or monitored.

Be sure to have the following information available before calling:

- Technical support registration number (if applicable)
- Product serial numbers
- Product model names and numbers
- Applicable error messages
- Operating system type and revision level
- Detailed, specific questions

HP Storage Web Site

The HP web site has the latest information on this product, as well as the latest drivers. Access the storage site at:

<http://www.hp.com/country/us/eng/prodserv/storage.html>. From this web site, select the appropriate product or solution.

HP NAS Services Web Site

The HP NAS Services site allows you to choose from convenient HP Care Pack Services packages or implement a custom support solution delivered by HP ProLiant Storage Server specialists and/or our certified service partners. For more information see us at

http://www.hp.com/hps/storage/ns_nas.html.

Introduction

HP Clustered File System and CFS-Linux provide scalability and high availability for the Network File System (NFS), which is commonly used on UNIX and Linux systems to share files remotely.

CFS-Linux Features

CFS-Linux provides the following features:

- **Scalable NFS client connectivity.** Over multiple NFS servers sharing the same filesystems, CFS-Linux supports a linearly increasing client connection load as similarly configured servers are added to the cluster. A 16-node cluster, serving the same filesystems via NFS, can support 16 times more NFS clients (with similar workloads and the same performance) than a single server. A price advantage is gained by using commodity-level Intel-based servers instead of larger SMP (4-way or 8-way) servers or larger proprietary filers to accommodate scaling client connectivity demands.
- **Scalable NFS performance.** With multiple NFS servers serving the same filesystems and with appropriate client balancing among the servers, CFS-Linux supports linearly increasing NFS performance. A 16-node NFS file-serving cluster provides nearly 16 times the performance results over a single NFS server for the same filesystems, up to the limit of the shared storage bandwidth.

- **Incremental NFS Server scalability.** New physical NFS servers can be added to an existing file-serving cluster to serve the same filesystems for a set of new clients without disturbing the connectivity and operation of other NFS clients and servers that may be sharing those filesystems.
- **High availability for NFS clients.** CFS-Linux supports continuous NFS client operation across a hardware or software failure that inhibits NFS service from a given cluster server. Because another cluster server inherits the same IP address used for the NFS client-server connections, the clients will continue NFS operation to that same IP address via a different physical server.

The server inheriting the IP address and associated NFS clients may also be an NFS server serving a pre-existing set of NFS clients. These pre-existing clients will not experience any interruption or delay when the new NFS clients transition to the server. (Only the transitioning subset of NFS clients, and not the pre-existing clients, are subject to the lock-recovery grace period and other transitional delays.) The transitioning subset of clients are shielded from transient errors that may be caused by the failure that induced the transition.

- **Cluster-wide NFS client lock coherency.** CFS-Linux supports mutual exclusion with file and byte-range locks used by multiple NFS clients connected to separate NFS servers exporting the same filesystem(s). Note, however, that file locks via NFS are disabled by default. Explicit administrative action is required to enable file locking via NFS. See “Using the NLM Protocol” on page 37 for more information.
- **NFS client lock re-acquisition on server failover.** When a group of NFS clients fail over to another NFS server, any file or byte-range locks held by those clients are released and then automatically re-acquired after the clients have successfully transitioned to another server.
- **Group migration of NFS clients.** The administrator can gracefully migrate an NFS service and the associated NFS clients from one server to another for maintenance or load-balancing purposes, without downtime or failure of any NFS client.

- **Cluster-wide consistent user authentication.** A user and NFS client may always be authenticated, by any NFS server in the cluster, as the same user and NFS client (assuming that an organizational authentication service such as LDAP has been deployed).
- **Cluster-wide, connection-oriented load balancing.** Through the use of DNS round-robin connection load balancing (or any external load balancer), NFS client connections mounting through a single common IP address (or DNS name) will be automatically and evenly distributed among the NFS servers in the cluster that are exporting the same filesystems. The DNS service may also be configured for high availability on the same file-serving cluster.

CFS-Linux Concepts and Definitions

CFS-Linux uses and manages the following objects to provide scalable and highly available file service across the cluster: Export Groups, export records, and Virtual NFS Services.

Export Groups

An Export Group is equivalent in form and function to the */etc/exports* file of a traditional NFS server. However, unlike a traditional NFS server, CFS-Linux supports multiple Export Groups (the equivalent of multiple independent */etc/exports* files) on a single node.

An Export Group is composed of a name (which must be unique among all other Export Groups in the cluster), a list of cluster nodes that will participate in the NFS service for the exports defined by this Export Group, and a list of Export Records describing the filesystems shared via NFS by the nodes listed in the Export Group.

An Export Group created on one node of the cluster is immediately available to all cluster nodes defined in the Export Group. A high-availability monitor is automatically associated with each of these nodes to monitor the health of the NFS service on the nodes and to initiate failover actions if the health of the NFS service degrades.

Export Records

An export record is equivalent, indeed exactly equal to, the individual records contained in the */etc/exports* file of a traditional NFS server. The format of the record and the options available are precisely the same in form and function.

Each record identifies a filesystem and directory sub-tree to be exported via NFS, the list of clients authorized to mount and access the specified filesystem sub-tree, and the options that modify how the filesystem sub-tree will be shared with clients (e.g., read-write or read-only; many other options exist). Free-form user comments can also be included with an export record.

The contents of an existing */etc/exports* file can be imported unchanged into a given Export Group. It is also possible to enter and edit individual export records directly on the Management Console. The filesystem sub-tree, shared via NFS, is referred to as an *export*.

Virtual NFS Services

A Virtual NFS Service is a virtual host (an HP Clustered File System VHOST) that is dedicated to NFS file service. Its primary purpose is to associate a “virtualized” IP address with an Export Group and to ensure, for high availability, that this virtualized IP address and the associated Export Group and NFS service are always available on one node of the cluster. A virtualized IP address is a portable IP address that can be moved from one node in the cluster to another when a failover action occurs.

A virtualized IP address, an associated Export Group, a “primary” cluster node, and a list of “backup” cluster nodes constitute a highly available Virtual NFS Service. NFS clients connected via the virtualized IP address will always remain connected to their Virtual NFS Service (with the same filesystem exports) regardless of any failover transitions, from one node to another, of the Virtual NFS Service.

When creating a Virtual NFS Service, the administrator specifies an unused network address that NFS clients will use to connect to the highly available Virtual NFS Service. This address must be on the same subnet as one of the networks already in use in the cluster.

The administrator then specifies both a primary node to host the Virtual NFS Service and an ordered set of backup nodes to host the Virtual NFS Service in case of failover. Finally, the administrator selects one Export Group to be associated with the Virtual NFS Service. This Export Group defines which exports will be available via NFS for this Virtual NFS Service.

It is possible to create multiple Virtual NFS Services and associate the same Export Group with each one. This enables “scale-out” NFS service for the same filesystems across multiple NFS servers in the cluster. It is further possible to set up a DNS round robin configuration such that a single IP address (or DNS name) can be used by all clients, yet the client connections will be evenly distributed among the several Virtual NFS Services running on separate nodes of the cluster. Other configurations are possible.

Supported NFS Version

CFS-Linux supports versions 2 and 3 of the NFS file sharing protocols based on UDP.

NOTE: TCP mode is not supported.

CFS-Linux Tested Configuration Limits

CFS-Linux has been tested to the following configuration limits:

- Up to 16 nodes in the cluster.
- 512 independent PSFS filesystems per cluster.
- 64 Virtual NFS Services per node with a maximum of 1 Export Group each.
- 64 Export Groups per node.
- 1024 export records per Export Group.

Configure Export Groups

An Export Group describes a set of PSFS filesystems to be exported. It also specifies the Virtual NFS Services that will provide virtual IP addresses that clients use to access those filesystems.

Overview

Export Records and Groups

The PSFS filesystems to be exported are described in export records, which are similar to the entries in an */etc/exports* file. They specify the directory to be made available for mounting by NFS clients, the clients that can mount the directory, and the permissions on the directory.

When you configure an Export Group, you will need to create an export record for each filesystem or directory to be exported. You can create the records either manually, or by importing an existing */etc/exports* file. You will also need to select one or more Virtual NFS Services to host the Export Group.

Virtual NFS Services

Each Export Group may be associated with one or more Virtual NFS Services. A Virtual NFS Service provides an IP address that NFS clients can use to access the PSFS filesystems exported by the associated primary server; in effect, it “virtualizes” the IP address associated with the NFS Service.

NFS clients access the filesystems through the Virtual NFS Service; they do not need to know which physical node is providing the service. Clients can access any of the filesystems specified in the Export Group by using the DNS name or IP address of the Virtual NFS Service.

When you create a Virtual NFS Service, you will need to assign it to a primary node and one or more backup nodes. A Virtual NFS Service can be associated with only one Export Group. However, you can assign the same Export Group to multiple Virtual NFS Services.

One or Many Export Groups?

In the typical configuration, a single Export Group associated with multiple Virtual NFS Services is recommended.

High Availability and Failover Support

CFS-Linux provides high availability for the exported PSFS filesystems. To ensure that the filesystems are always available, CFS-Linux provides a high-availability monitor for each Export Group. On each server configured for the Export Group, this monitor periodically checks the exported filesystems and the NFS daemon to determine whether they can be accessed externally.

If the server to which NFS clients are connected should fail or lose access to the PSFS filesystem (for example, because of a SAN problem), the monitor probe will report the failure. The Virtual NFS Services using that server as their primary are then failed over (redirected) to a backup server. Failover is transparent to the NFS clients; their access to the exported filesystems continues after a brief pause.

Caveats Regarding Export Groups

Before creating an Export Group, you should be aware of the following:

- Although you can configure multiple Export Group files, with each group having its own set of export records, all of the export records are ultimately merged into a single */etc/exports* file that is automatically installed on each node in the cluster.

Because an export can be included in more than one Export Group, the */etc/exports* file can contain multiple records referring to the same filesystem path to be exported. If these duplicate records specify different export options, it is indeterminate which options the NFS service will assume. It is therefore recommended that any given filesystem subtree be exported by at most one Export Group.

- Although an Export Group can specify that only certain nodes can export its filesystems, the contents of the Export Group are available via the */etc/exports* file on every node in the cluster. This means that users can access the exported filesystems from nodes where the Export Group is not defined by connecting via the physical address of the node. However, such access is not highly available because there is no Virtual NFS Service associated with those connections. This is not a security concern because the entire cluster should be considered as a single security domain (for this and other reasons). The administrator can track which nodes are exporting *highly available* NFS services by watching the Export Group high-availability monitors that are active on those nodes.

Mount PSFS Filesystems

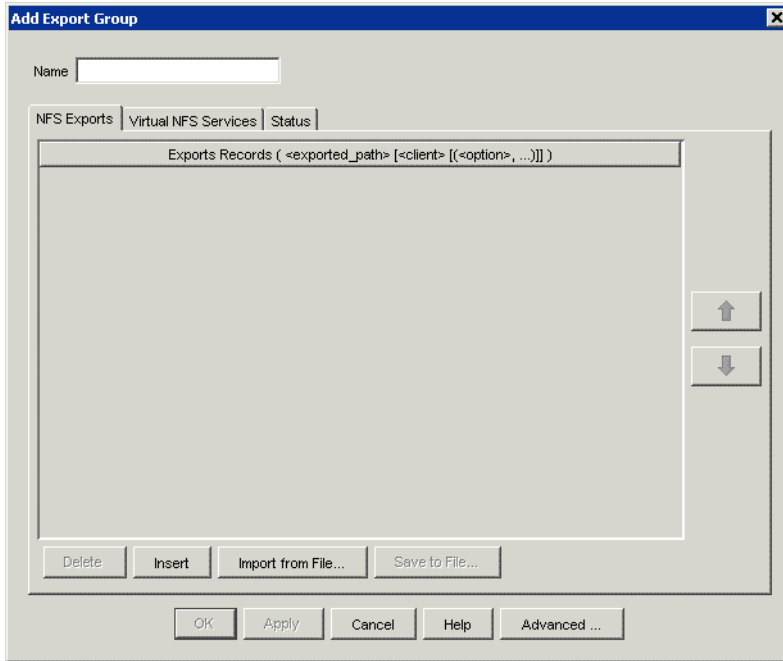
Before configuring Export Groups, ensure that the filesystems to be exported are on PSFS filesystems and are mounted on all the pertinent nodes with the “persist” mount option.

We also recommend that you use the “sync” mount option for safety reasons. Although the “async” option provides better NFS write performance, as it allows writes to be acknowledged before being committed on disk, it gives less coherency with regard to on-disk contents in the cluster. Consequently, server crashes can lead to silent-data-loss; the NFS client may never be aware of a given server crashing because CFS-Linux provides for seamless fail-over.

See “Configure PSFS Filesystems,” in the *HP Clustered File System Administration Guide* for details about creating and mounting PSFS filesystems.

Add an Export Group

To create an Export Group, select Solution Packs > MxFS > Add Export Group. You can also right-click on an existing Virtual NFS Service that is not currently assigned to an Export Group and select Add Export Group.



The Add Export Groups window has three tabs. The NFS Exports and Virtual NFS Services tabs are used to create the Export Group. The third tab, Status, reports the current state of the Export Group.

At the **Name** prompt, type a name for this Export Group. The name must be unique among all of the Export Groups in the cluster and cannot include spaces.

Then complete the NFS Exports and Virtual NFS Services tabs.

Click Apply to save changes either since the last use of Apply or since the dialog was opened. When you click Apply, the dialog remains open. Clicking OK also saves changes, but closes the dialog.

NFS Exports Tab

This tab is used to configure the export records for the Export Group. An export record specifies a PSFS filesystem or directory to be exported, the NFS clients allowed to mount it, and the export options for the client.

You can type the export records directly on the form, optionally using the Export Record Details window to construct the record, or you can import an existing exports file.

Import an Existing Exports File

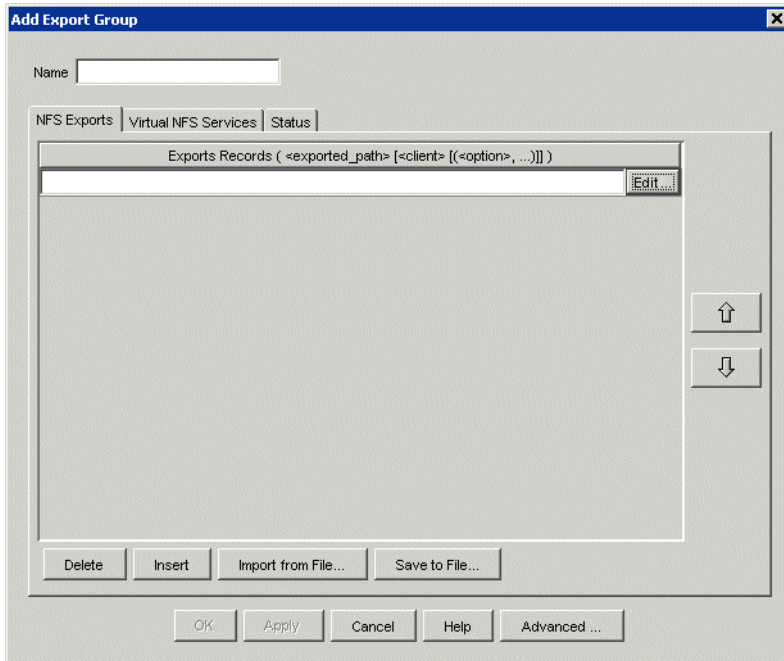
To import an existing file, click on Import from File and then select the file, which must be on the local server that is running the Management Console. The default Import location is the user's home directory. When you select the appropriate file, you will be asked whether you want the records in the file to replace any existing records in the Export Group or to be appended to the existing records.

Comments are preserved from the imported file; however, any blank lines are removed. The continuation character (\) is also removed as this character is not allowed in an export record.

NOTE: If an */etc/exports* file existed on the server when CFS-Linux was installed, the file will have been preserved as */etc/exports.pre_mxfs*.

Create Export Records

To add export records directly to the NFS Exports tab, click on Insert. A row will then open up on the NFS Exports tab.



You can type the record information directly on the row, or you can click the Edit button at the end of the row to display the Export Record Details dialog, which simplifies constructing the record.

An export record uses the following format, which is as documented in the Linux **exports(5)** man page:

```
<exported_path> [<client>] [( <option>, ...)]
```

The *<exported_path>* must start with a slash (/) and must include at least one directory name (for example, */data1* is valid but */* is invalid). If you do not specify a client, all clients will be allowed access. An asterisk (*) is also equivalent to specifying all clients. The options are as specified in the Linux **exports(5)** man page. All options are supported, and comments are allowed.

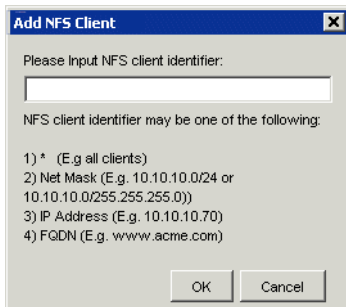
If you click the Edit button, the Export Record Details dialog appears. This dialog includes a Preview line at the bottom that shows the export record as you construct it. Following is an example of a completed export record.

The dialog box is titled "Export Record Details". It contains the following fields and controls:

- Exported Path:** A text field containing "/data1".
- Client Names:** A list box containing "10.10.10.70". Above the list are "Add.." and "Delete" buttons.
- Options:** A group box containing several radio button options:
 - ☒ no_all_squash ☐ all_squash
 - ☒ nocrossmnt ☐ crossmnt
 - ☒ wdelay ☐ no_wdelay
 - ☒ hide ☐ nohide
 - ☒ auth_nlm ☐ no_auth_nlm
 - ☐ ro ☒ rw
 - ☒ root_squash ☐ no_root_squash
 - ☒ secure ☐ insecure
 - ☒ subtree_check ☐ no_subtree_check
 - ☒ sync ☐ async
- anongid:** A text field containing "-2".
- anonuid:** A text field containing "-2".
- Comment:** A text field containing "client70".
- Preview:** A text area showing the resulting export line: `/data1 * 10.10.10.70(rw) #client70`.
- Buttons:** "OK" and "Cancel" buttons at the bottom right.

Exported Path. Specify the filesystem or directory to be exported.

Client Name: Click Add to specify a client to be allowed to mount the exported filesystem. Then, on the Add NFS Client dialog, identify the client by either its Netmask, IP address, or FQDN. Optionally, you can enter an asterisk (*) to specify all clients.



Options. Select any options that apply to this record. The default options are selected initially and do not appear on the Preview line.

After you have created the export records, you can use the arrow buttons on the NFS Exports tab to reorder the entries as you prefer them.

NOTE: When the Export Group is created, CFS-Linux validates each export record and reports any syntax errors.

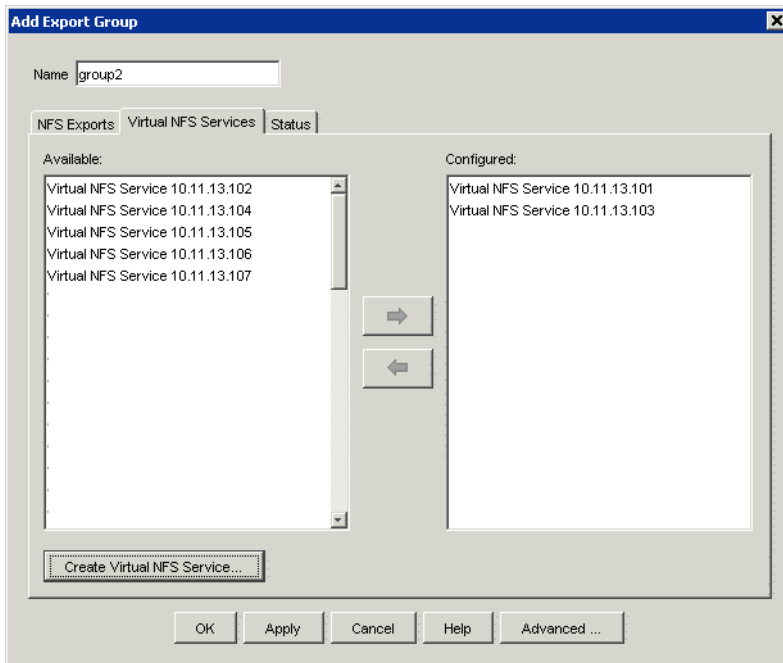
Save Export Records to a Local File

The Save to File option on the NFS Exports tab allows you to save to a local file the export records for a particular Export Group. The file you specify will be created if it does not already exist. If the file already exists, you will be asked to confirm that the file should be replaced. You can use this option if you want to manually examine or edit the records you have created and later re-import them.

Virtual NFS Services Tab

On this tab, assign the Export Group to one or more Virtual NFS Services. (A Virtual NFS Service can be associated with only one Export Group; however, an Export Group can be assigned to multiple Virtual NFS Services.)

The Available column on the Virtual NFS Services tab lists the Virtual NFS Services that are not currently associated with an Export Group. Use the arrow buttons to move the appropriate Virtual NFS Services from the Available column to the Configured column.



If the appropriate Virtual NFS Service does not currently exist, you can create it by clicking the “Create Virtual NFS Service” button. (See “Add a Virtual NFS Service” on page 30 for more information.)

You can now click Apply or OK to create the Export Group, or you can configure the advanced options as described under “Advanced Options for Export Groups” on page 17. (The advanced options are for the high-availability monitor associated with the Export Group.)

When you click Apply or OK, CFS-Linux validates each export record and reports any syntax errors.

To create an Export Group from the command line, type the following command:

```
mx exportgroup add [<advanced-options>][--exports <exports_file>
--vnfs <vnfs1>,<vnfs2>,...] <exportgroup_name>
```

Export Group High-Availability Monitor

After an Export Group is created, CFS-Linux starts a high-availability monitor on each node associated with the Export Group. This monitor is active on those nodes and watches the NFS processes running on them.

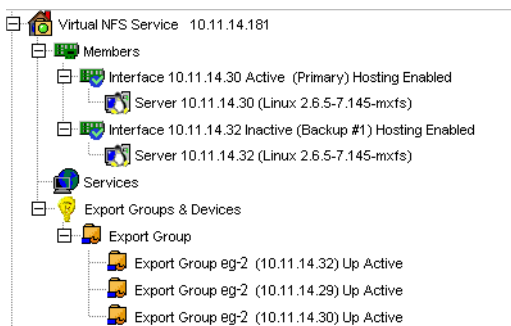
View the High-Availability Monitor

The high-availability monitor, which is labeled “Export Group <name>,” looks like this on the Servers tab of the Management Console.



In this example, Virtual NFS Service 10.11.14.181 is active on node 10.11.14.30. The Virtual NFS Service is exporting Export Group eg-2. The monitor associated with eg-2 (labeled Export Group eg-2) is up and active on node 10.11.14.30.

The high-availability monitor also appears on the Virtual Hosts tab. This example for Virtual NFS Service 10.11.14.181 shows the network interfaces (and nodes) on which the Virtual NFS Service is configured. It also shows the configuration of Export Group eg-2. Note that this Export Group is assigned to two Virtual NFS Services: Virtual NFS Service 10.11.14.181 as shown below, and the Virtual NFS Service associated with network interface 10.11.14.29.



How the High-Availability Monitor Works

When configuring the high-availability monitor associated with an Export Group, you should be aware of the actions taken by the monitor. During each probe cycle, the monitor performs a series of checks:

1. The monitor first checks basic NFS Server health by issuing a NULL RPC call to the NFS Server on the local node. If this call fails, the Export Group is considered to be DOWN.
2. The monitor next checks the general health of the CFS-Linux high-availability service by checking for critical CFS-Linux processes. If any of these processes are not running, the Export Group is considered to be DOWN. (The processes may not be running because the node is still initializing or shutting down.)
3. The monitor then checks that each exported path in the Export Group is available and is mounted on the PSFS filesystem. The monitor does these checks in timed background threads. If any exported path cannot be verified in one-half of the probe's timeout value (or five seconds, whichever is greater), the Export Group is considered to be DOWN, as the verifying operations are simple and should reasonably be completed in this time period.

Because the third check takes a linear amount of time based on the number of exported paths in the Export Group, it is important to set the probe timeout value to something larger than the default value of 15 seconds if the Export Group contains more than a few exported paths. If the high-availability monitor cannot obtain reasonable responsiveness from the system for exported paths, neither can NFS clients. This check verifies the practical availability of an exported path as well as its literal availability.

In the third check, the monitor only issues a warning if a path does not exist (instead of considering the Export Group to be DOWN as it does in the situations described above). This method avoids problems that can arise if an exported path is deleted from a shared filesystem.

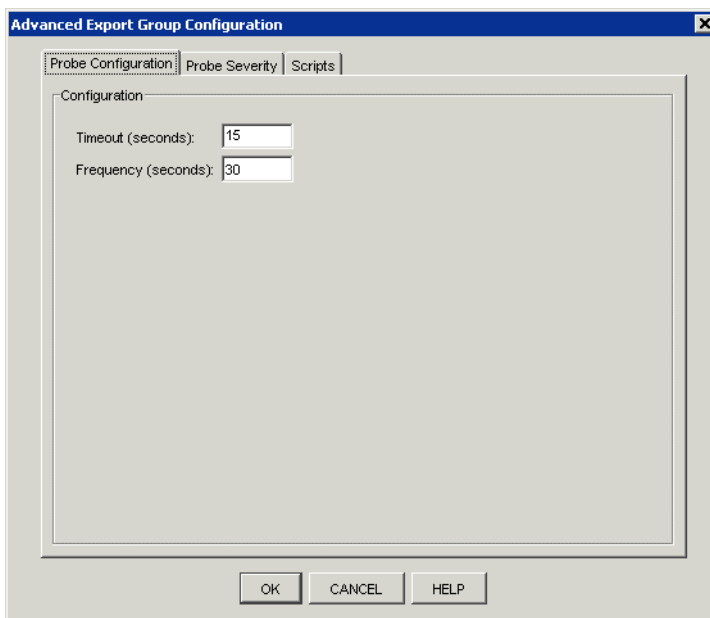
Because the purpose of the monitor is to determine whether the local node is healthy enough to host the NFS service, rather than considering the entire Export Group to be DOWN on a cluster-wide basis, the monitor will warn than it cannot find the export path. A deleted export path should be either recreated or removed from the Export Group at the administrator's earliest convenience.

Advanced Options for Export Groups

The advanced options allow you to tailor the configuration of the high-availability monitor associated with an Export Group.

Probe Configuration

The Probe Configuration tab can be used to configure the timeout and frequency for the probe initiated by the high-availability monitor.



Timeout: The maximum amount of time that CFS-Linux will wait for a probe to complete. The default is 15 seconds. This value should be adequate for most environments.

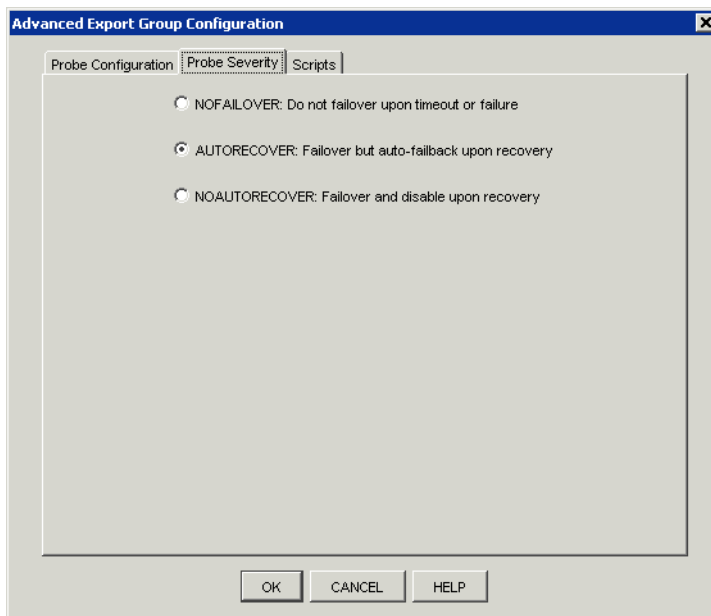
However, if `nsfd` is extremely busy (as in HPC environments), you may need to increase the timeout to prevent false failovers of the Virtual NFS Service. If you see false failovers, increase the timeout in 30-second increments until the false failovers stop.

Also note that the timeout should be increased for Export Groups containing more than a few exported paths. A greater timeout value ensures that the probe has adequate time to verify that each exported path in the Export Group is available and is mounted on the PSFS filesystem.

Frequency: The interval of time, in seconds, at which the monitor probes the NFS Service. The default is 30 seconds.

Probe Severity

The Probe Severity tab lets you specify the failover behavior of the high-availability monitor.



The Probe Severity setting works with the Virtual NFS Service policy (either AUTOFAILBACK or NOFAILBACK) to determine what happens when a monitor probe fails.

The default policies (AUTOFAILBACK for the Virtual NFS Service and AUTORECOVER for the high-availability monitor) cause HP Clustered File System to fail over the associated Virtual NFS Services to a backup network interface on another node when the monitor probe fails. When the NFS service is restored on the original node, HP Clustered File System fails back the Virtual NFS Services to the network interfaces on the original node.

You can use the Probe Severity attribute to change this behavior. There are three settings for Probe Severity: NOFAILOVER, AUTORECOVER, and NOAUTORECOVER.

NOFAILOVER. The Virtual NFS Services do not fail over when the monitor probe is DOWN. Administrative intervention is required to fail over the Virtual NFS Services.

AUTORECOVER. This is the default. The Virtual NFS Services fail over when a monitor probe fails. When the NFS service is recovered on the original node, failback occurs according to the failback policy for the Virtual NFS Services.

NOAUTORECOVER. The Virtual NFS Services fail over when a monitor probe fails and the monitor is disabled on the original node, preventing automatic failback. When the monitor is reenabled, failback occurs according to the failback policy for the Virtual NFS Services.

The NOAUTORECOVER option is useful when integrating CFS-Linux with a custom application where certain application-specific actions must be taken before failback can occur.

To set the Probe Severity from the command line, use this option:

```
--probeSeverity nofailover|autorecover|noautorecover
```


Scripts

The Scripts tab lets you configure custom Recovery, Start, and Stop scripts for the high-availability monitor.

The image shows a screenshot of the 'Advanced Export Group Configuration' dialog box, specifically the 'Scripts' tab. The dialog has three tabs: 'Probe Configuration', 'Probe Severity', and 'Scripts'. The 'Scripts' tab is active. It contains a table with two columns: 'Script pathname' and 'Timeout (seconds)'. There are three rows for 'Recovery:', 'Start:', and 'Stop:'. Below the table, there are two sections: 'Event Severity' and 'Script Ordering'. The 'Event Severity' section has two radio buttons: 'CONSIDER events in making failover decisions' (selected) and 'IGNORE events in making failover decisions'. Below it is a note: 'Note: Events are configuration errors and start/stop script timeouts/failures.' The 'Script Ordering' section has two radio buttons: 'SERIAL: During matrix transition, wait for stop scripts to succeed before running start script' (selected) and 'PARALLEL: Run start script while stop scripts are running'. Below it is a note: 'Note: SERIAL will consider events and takes precedence over event severity.' At the bottom of the dialog are three buttons: 'OK', 'CANCEL', and 'HELP'.

Script pathname	Timeout (seconds)
Recovery:	
Start:	
Stop:	

Event Severity

☒ CONSIDER events in making failover decisions

☐ IGNORE events in making failover decisions

Note: Events are configuration errors and start/stop script timeouts/failures.

Script Ordering

☒ SERIAL: During matrix transition, wait for stop scripts to succeed before running start script

☐ PARALLEL: Run start script while stop scripts are running

Note: SERIAL will consider events and takes precedence over event severity.

OK CANCEL HELP

Monitors can optionally be configured with scripts that are run at various points during cluster operation. The script types are as follows:

Recovery script. Runs after a monitor probe failure is detected, in an attempt to restore the NFS service.

Start script. Runs as the NFS service is becoming active on a server.

Stop script. Runs as the NFS service is becoming inactive on a server.

When a monitor is instantiated, HP Clustered File System chooses the best node to make the service active. The Start script is run on this node. The Stop script is run on all other nodes configured for the monitor to ensure that the service is not active on those nodes.

Start scripts must be robust enough to run when the service is already started, without considering this to be an error.

Similarly, Stop scripts must be robust enough to run when the service is already stopped, without considering this to be an error. In both of these cases, the script should exit with a zero exit status.

This behavior is necessary because the Start and Stop scripts are run to establish the desired start/stop activity, even though the service may actually have been started by something other than CFS-Linux.

The Start and Stop scripts must also handle recovery from events that may cause them to run unsuccessfully. For example, if the system encounters a problem, the script will fail and exit non-zero. The service could then become active on another node, causing the Stop script to run on the original node even though the Start script did not complete successfully.

To configure scripts from the command line, use these options:

```
--recoveryScript <script>  
--recoveryTimeout <seconds>  
--startScript <script>  
--startTimeout <seconds>  
--stopScript <script>  
--stopTimeout <seconds>
```

Event Severity

By default, CFS-Linux treats the failure or timeout of a Start or Stop script as a failure of the associated service and may initiate failover of the associated Virtual NFS Services. Configuration errors can also cause this behavior.

Such a failure or timeout creates an event associated with the monitor on the node where the failure or timeout occurred. You can view these events on the Management Console and clear them after you have fixed the problems that caused them.

You can configure the failover behavior with the Event Severity attribute. There are two settings:

CONSIDER. This is the default value. Events are considered during failover decisions.

IGNORE. Events are ignored and Start or Stop script failures will not cause failover. This is useful when the action performed by the Start and Stop scripts is not critical, but is important enough that you want to keep a record of it.

To configure event severity from the command line, use this option:

```
--scriptSeverity consider|ignore
```

Script Ordering

Script ordering determines the order in which the Start and Stop scripts are run when a Virtual NFS Service moves from one node to another. If you do not configure Start and Stop scripts for an Export Group, the script ordering configuration has no effect.

There are two settings for script ordering.

SERIAL. This is the default setting. When a Virtual NFS Service moves from one node to another, HP Clustered File System enforces the following strict ordering sequence for running Start and Stop scripts:

1. HP Clustered File System runs the Stop script on all nodes where the Virtual NFS Service should be inactive.
2. HP Clustered File System waits for all Stop scripts to complete.
3. HP Clustered File System runs the Start script on the node where the Virtual NFS Service is becoming active.

PARALLEL. HP Clustered File System does not enforce the strict ordering sequence for Stop and Start scripts. The scripts run in parallel across the cluster as a Virtual NFS Service is in transition.

The **PARALLEL** configuration can speed up failover time for services that do not depend on strict ordering of Start and Stop scripts. Assuming that it is safe to run the scripts in parallel (which depends on your application), this setting can also increase the chances of a successful failover because HP Clustered File System does not have to wait for the failing node to finish running its Stop script.

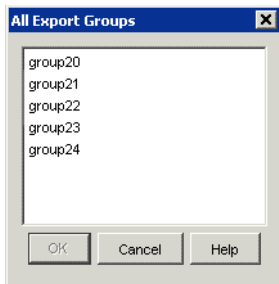
To configure script ordering from the command line, use this option:

```
--ordering serial|parallel
```

View Export Group Properties

To view all of the Export Groups configured with CFS-Linux, select Solution Packs > MxFS > View All Export Groups, or select an Export Group on the Applications tab, right-click, and select View All Export Groups.

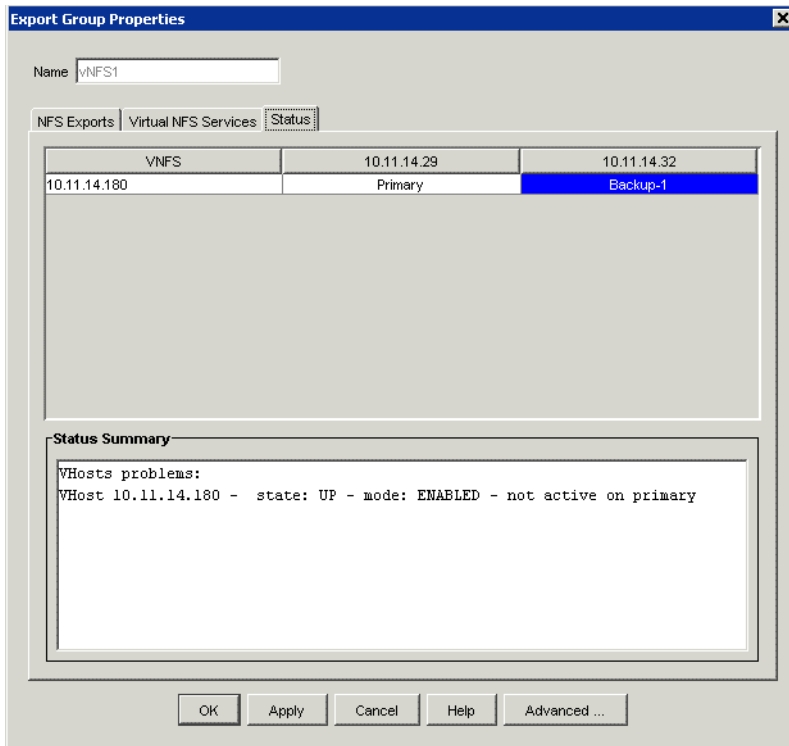
The All Export Groups window then appears. This window lists the names of all Export Groups that have been configured.



To see the properties for a particular Export Group, highlight that Export Group and click OK. The Export Group Properties window then appears. This window shows the Export Group as it is currently configured.

The Status tab on the Export Group Properties window shows the network interfaces on which the associated Virtual NFS Services are configured, including their status as a primary or backup for the Export Group. The primary interface for the Virtual NFS Service appears in blue.

The Status Summary at the bottom of the window reports any errors.



To view status from the command line, use the following command:

```
mx exportgroup status [--up|--down] [--enabled|--disabled]  

[--primary|--backup] [--active|--inactive] [<exportgroup_name>  

...]
```

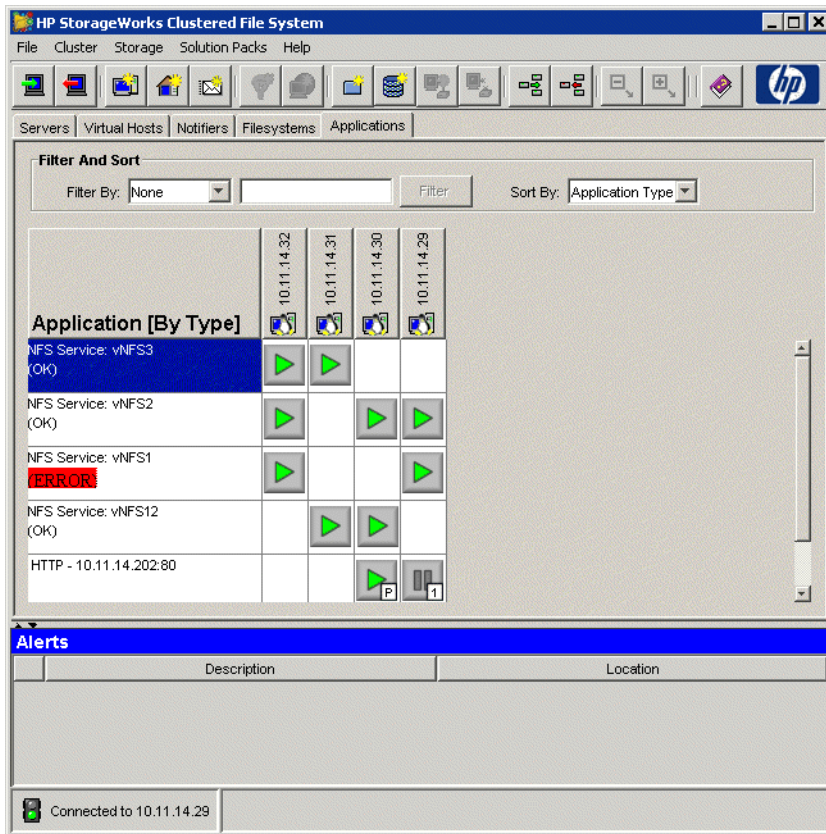
Modify an Export Group

To modify an existing Export Group, go to the Export Group Properties window, as described above. To change an export record, go to the NFS Exports tab. Select the appropriate record, and either make your changes directly on the record, or click the Edit button to display the Export Record Details dialog. You can also use the Virtual NFS Services tab to change the Virtual NFS Services assigned to the Export Group.

Changes to an Export Group become effective almost immediately and can affect clients.

Manage CFS-Linux from the Applications Tab

The Applications tab on the Management Console lists each Export Group and the servers on which the associated Virtual NFS Services are configured.



The Application column reports the status of each Export Group. The status will be one of the following:

- OK. The Export Group is healthy. All of its high-availability monitors are currently running and active on the associated nodes.

All of the Virtual NFS Services associated with the Export Group are currently running and active on their primary interfaces.

- **WARNING.** The Export Group is available, but currently fails to meet one or more of the health conditions described under OK, above. For example, a Virtual NFS Service may have failed over to a backup interface, and administrative action may be necessary.
- **ERROR.** The Export Group currently fails to meet all of the health conditions described under OK, above. The Export Group is no longer available, or one or more Virtual NFS Services may be down and not running on any node.

When you right-click on an Export Group, the following options are available:

- View properties for this Export Group, including status.
- Delete this Export Group.
- Add a new Export Group.
- Rehost the Export Group.
- Enable or disable this Export Group on all servers.
- View all Export Groups.

You can also administer Export Groups at the server level. In the server column, select the cell corresponding to the Export Group. Right-click to display the following options:

- View properties for this Export Group, including status.
- Delete this Export Group.
- Enable or disable the Export Group on this server.
- Rehost the Export Group.
- View or clear the last error for this Export Group on this server.
- Add a new Export Group.
- View all Export Groups.

Other Export Group Procedures

Disable an Export Group

Select the high-availability monitor associated with the Export Group on the Servers or Application tab, right-click, and select Disable.

To disable the high-availability monitor from the command line, use this command. If no servers are specified, the action takes place on all servers.

```
mx exportgroup disable <exportgroup_name>  
[ALL_SERVERS] | [<server> ...]
```

Enable an Export Group

Select the high-availability monitor associated with the Export Group on the Servers or Application tab, right-click, and select Enable.

To enable the high-availability monitor from the command line, use this command. If no servers are specified, the action takes place on all servers.

```
mx exportgroup enable <exportgroup_name>  
[ALL_SERVERS] | [<server> ...]
```

Clear an Error Associated with an Export Group

To clear a error from a high-availability monitor associated with an Export Group, select that monitor on the Servers or Application tab, right-click, and select Clear Last Event.

To clear the error from the command line, use this command:

```
mx exportgroup clear <exportgroup_name> <server> ...
```

Delete an Export Group

To delete an Export Group and the high-availability monitor associated with it, select the high-availability monitor on the Servers or Applications tab, right-click, and select Delete.

To delete the Export Group from the command line, use this command:

```
mx exportgroup delete <exportgroup_name> ...
```

Configure Virtual NFS Services

A Virtual NFS Service exports the PSFS filesystems specified in its associated Export Group. Clients access the filesystems via the hostname/IP address of the Virtual NFS Service instead of using the hostname/IP address of the physical node.

To create a Virtual NFS Service, you will need to specify the nodes on which the Virtual NFS Service should be configured. Optionally, you can also specify the Export Group that should be exported by the Virtual NFS Service. (You can also associate a Virtual NFS Service with an Export Group when you create the Export Group.)

Overview

Sample Configurations

Following are some examples of configurations using CFS-Linux to provide NFS services.

Active-Active Failover Configuration

In an active-active configuration, all nodes in the cluster may act as a primary for one or more Virtual NFS Services and as a backup for other Virtual NFS Services. If an Export Group monitor reports a failure on a particular node, CFS-Linux may fail over Virtual NFS Services from that node to a backup node. Clients then continue to access the Virtual NFS Services via the same IP address from the backup node without downtime or disconnection.

Because the files associated with the NFS Service are located on a PSFS cluster filesystem, the backup node can immediately access the same filesystem data and continue with any I/O operations in progress.

This type of configuration provides the best possible utilization of resources in the cluster, while also preserving high availability.

Active-Passive Failover Configuration

In an active-passive configuration, one or more nodes act as a backup for all of the Virtual NFS Services running on the other nodes. The passive nodes are running NFS, but are not the primary for any Virtual NFS Services. If an Export Group monitor reports a failure on an active node, the associated Virtual NFS Service will fail over to a passive node, which will begin hosting the Export Group. This type of configuration is useful when the NFS service workload is reaching the capacity of the active servers and they will not be able to absorb the workload from another server if it should fail over.

Guidelines for Creating Virtual NFS Services

When creating Virtual NFS Services, follow these guidelines:

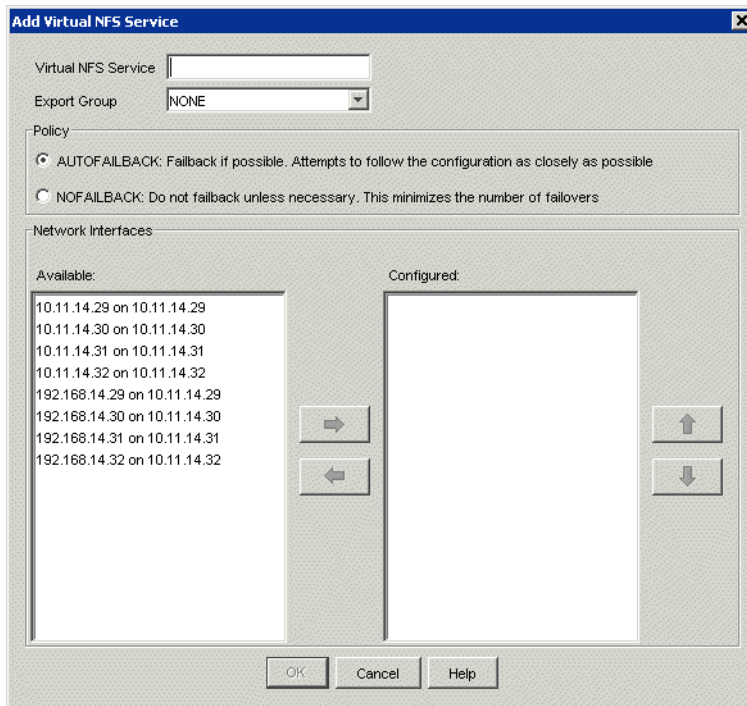
- When planning the Virtual NFS Services needed for your cluster, first determine the network services that will be available to your clients. Then determine the IP addresses for those services. You will need to create a Virtual NFS Service for each IP address.
- Choose hostnames that differ from your actual server names. Virtual NFS Services are independent of specific servers, and their names should be independent as well.
- Use an IP address that is on the same subnet as the network interfaces where it will be configured.
- Update the DNS name service or the */etc/hosts* file with the virtual hostnames and IP addresses. (For improved performance, the Management Console caches hostname lookups. If your DNS changes, you may need to restart the console so that it will reflect the new hostname.)
- Do not use **ifconfig** or another tool to configure the IP address in the operating system or a system startup script.

HP Clustered File System configures the operating system appropriately to support the Virtual NFS Service.

- Virtual NFS Services share a common name-space with other HP Clustered File System virtual hosts. You cannot create a virtual host and a Virtual NFS Service having the same network address.
- After creating Virtual NFS Services, you will need to configure your applications to recognize them.

Add a Virtual NFS Service

To add a virtual NFS Service, select Solution Packs > MxFS > Add Virtual NFS Service.



The "Add Virtual NFS Service" dialog box contains the following fields and sections:

- Virtual NFS Service:** A text input field.
- Export Group:** A dropdown menu currently showing "NONE".
- Policy:** Two radio button options:
 - ☒ **AUTOFAILBACK:** Failback if possible. Attempts to follow the configuration as closely as possible
 - ☐ **NOFAILBACK:** Do not failback unless necessary. This minimizes the number of failovers
- Network Interfaces:** A section with two panes:
 - Available:** A list of network interfaces:
 - 10.11.14.29 on 10.11.14.29
 - 10.11.14.30 on 10.11.14.30
 - 10.11.14.31 on 10.11.14.31
 - 10.11.14.32 on 10.11.14.32
 - 192.168.14.29 on 10.11.14.29
 - 192.168.14.30 on 10.11.14.30
 - 192.168.14.31 on 10.11.14.31
 - 192.168.14.32 on 10.11.14.32
 - Configured:** An empty list box.

Navigation buttons are located between the panes: a right arrow (→) to move an interface from Available to Configured, and a left arrow (←) to move it back. On the right side of the Configured pane, there are up (↑) and down (↓) arrow buttons for reordering.

At the bottom of the dialog are three buttons: **OK**, **Cancel**, and **Help**.

Virtual NFS Service: Enter a hostname or IP address for this Virtual NFS Service.

Export Group: Select the Export Group that will be exported by this Virtual NFS Service. If the Export Group has not yet been created, select NONE.

Policy: The policy determines the failback action that the Virtual NFS Service will take following a failover to a backup node.

- **AUTOFAILBACK.** This policy is intended to return the Virtual NFS Service to its original configuration, or as close to it as possible. After the Virtual NFS Service fails over to a backup node, HP Clustered File System watches the health of the nodes that are higher in the list of servers configured for that Virtual NFS Service. When the health of one of these nodes is *equal to or greater* than the backup node where the Virtual NFS Service currently resides, the Virtual NFS Service will automatically attempt to fail over to that node.
- **NOFAILBACK.** This policy is intended to minimize failovers. The Virtual NFS Service remains active on the backup node until a “healthier” node becomes available, at which point the Virtual NFS Service fails over to that node. (On a “healthier” node, *more* of the services associated with the Virtual NFS Service will be up than on the node currently hosting the Virtual NFS Service.)

Network Interfaces Available/Configured: Move the interfaces on which the Virtual NFS Service should be configured from the Available column to the Configured column. The first interface that you select is the primary interface. The other interfaces that you select are backups. You can use the up and down arrows to reorder the interfaces on the Configured column.

To create a Virtual NFS Service from the command line, use this command:

```
mx vnfs add [--exportgroup <exportgroup_name>|NONE]
[--policy autofailback|nofailback] <vnfs> <network_interface>
...
```

You can view the configuration and status of Virtual NFS Services and the Export Groups associated with them on the Management Console. See “Export Group High-Availability Monitor” on page 15 and also “Manage CFS-Linux from the Applications Tab” on page 25.

To view Virtual NFS Services from the command line, use this command:

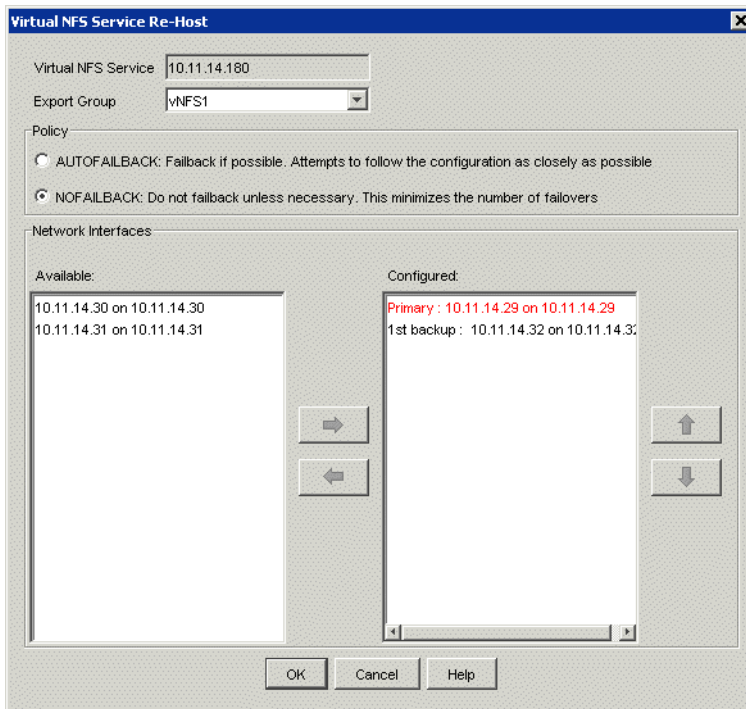
```
mx vnfs status [--up|--down] [--enabled|--disabled]  
[--primary|--backup] [--active|--inactive] [<vnfs> ...]
```

Migrate a Virtual NFS Service

The Re-Host Virtual NFS Service option allows you to move a Virtual NFS Service to another node. For example, you might want to move the Virtual NFS Service from the primary node before taking that node down for maintenance. You can use this option in the following ways.

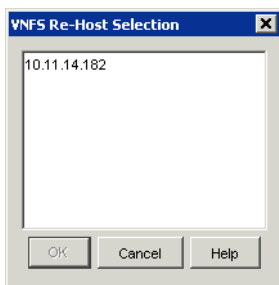
From the Servers or Virtual Hosts Tab

Select the Virtual NFS Service on either the Servers or Virtual Hosts tab, right-click, and select Re-Host Virtual NFS Service. Then, on the Virtual NFS Service Re-Host window, use the arrows to move the Virtual NFS Service to another node. Use the up and down arrows to reorder the interfaces associated with the primary and backup nodes. To change the set of nodes used as a primary or backup for the Virtual NFS Service, move the appropriate interfaces to the Available or Configured columns.



From the Application Tab

On the Applications tab, select the Export Group associated with the Virtual NFS Service, right-click, and then select Export Group Re-Host. The VNFS Re-Host Selection window then appears.



Choose the hostname or IP address associated with the Virtual NFS Service that you want to rehost. (Depending on your configuration, the Export Group may be associated with more than one Virtual NFS Service.) When the Virtual NFS Service Re-Host window appears, you can modify the nodes used as the primary or a backup for the Virtual NFS Service.

From the Command Line

Issue the following command, where *<vnfs>* is the Virtual NFS Service to be rehosted. You will need to specify all network interfaces on the which the Virtual NFS Service should be configured (the primary and all backups).

```
mx vnfs move <vnfs> <networkinterface> ...
```

Other Virtual NFS Service Procedures

Disable a Virtual NFS Service

Select the Virtual NFS Service on the Virtual Hosts or Applications tab, right-click, and select Disable.

To disable the Virtual NFS Service from the command line, use this command:

```
mx vnfs disable <vnfs>
```

Enable a Virtual NFS Service

Select the Virtual NFS Service on the Virtual Hosts or Applications tab, right-click, and select Enable.

To enable the Virtual NFS Service from the command line, use this command:

```
mx vnfs enable <exportgroup_name> <vnfs>
```

Delete a Virtual NFS Service

Select the Virtual NFS Service on the Virtual Hosts or Applications tab, right-click, and select Delete.

To delete the Virtual NFS Service from the command line, use this command:

```
mx vnfs delete <vnfs> ...
```


Other Configuration Information

This chapter provides information about the following:

- NFS clients
- Using the NLM protocol

NFS Clients

After CFS-Linux is configured, your NFS clients can begin accessing the exported PSFS filesystems.

Timeout Configuration

It is recommended that NFS clients have a minimum timeout value of 120 seconds. NFS failovers typically take much less time, but in a worst-case scenario may approach 120 seconds.

Client Mounts

To access the shared data on PSFS filesystems, clients simply mount the exported PSFS filesystems.

```
# mkdir /mnt/data1
# mount -t nfs -o udp 99.10.210.100:/mnt/psfs/nfsdata1
/mnt/data1
```

The following command verifies that clients on the 99.10.210.100 network can now access the shared data:

```
# ls -l /mnt/data1
total 1
drwxr-xr-x  2 root    root      48 Mar 24 00:32 dir1/
-rw-r--r--  1 root    root        0 Mar 24 00:32 file1
-rw-r--r--  1 root    root        0 Mar 26 12:29 file2
-rw-r--r--  1 root    root        0 Mar 24 00:32 file3
```

Using the NLM Protocol

NLM is the locking protocol used by NFS. By default, it is disabled when CFS-Linux is installed. If necessary, NLM can be enabled; however, you should be aware of the following caveat:

- File locks granted by the NFS server are cluster-coherent. When a failover occurs, the locks are released by the original server and the client automatically reclaims them on the new server (the backup node). However, during the period after the lock is released, another client or application may compete for and win the lock. Some NFS clients will return an error to the client applications if the lock cannot be reclaimed. Other clients (for example, the Linux 2.6 NFS client) will not return any error. If no error is returned by the client, the application may proceed under the false assumption that the lock has been granted. Data corruption may be the result.

To prevent this situation, locking should be enabled only if your clients are partitioned so that all clients needing a particular lock are using the same Virtual NFS Service IP address. If a failover occurs, all of the clients will lose their locks. They can then reclaim those same locks on the new node without conflicts from outside clients.

The **mxnlmconfig** command is used to enable or disable NLM locking in the cluster.

NOTE: The change takes place immediately and may affect clients.

The **mxnlmconfig** command has this syntax:

```
/opt/hpcfcs/bin/mxnlmconfig -q|-e|-d|-?
```

The options are as follows:

- q Show the current status of NLM locking in the cluster (either enabled or disabled).
- e Enable NLM locking in the cluster. No reboot is necessary; the change is effective almost immediately.
- d Disable NLM locking in the cluster. No reboot is necessary; the change is effective almost immediately.
- ? Display a syntax message.

Index

A

- Applications tab
 - Export Group, manage 25
 - Virtual NFS Service, move 33

C

- configuration
 - active-active 28
 - active-passive 29
 - examples 28
 - limits 5

E

- errors, status summary 23
- event, clear 21
- Export Group
 - add 9
 - advanced options
 - event severity 21
 - probe configuration 17
 - probe severity 18
 - script ordering 22
 - scripts, custom 20
 - Applications tab, manage from 25
 - assign to Virtual NFS Service 13
 - caveats 7
 - clear errors 27
 - defined 6
 - delete 27
 - disable 27
 - enable 27
 - object definition 3
 - status 24
 - view properties 23
- export records
 - clients, add 12

- configure 10
- defined 6
- Export Record Details dialog 12
- format 11
- import from file 10
- object definition 4
- options 11
- save 13
- exports file
 - access 7
 - import 10

G

- getting help v

H

- HP
 - storage web site v
 - technical support v

I

- installation
 - NFS version 5
 - operating system version 5

M

- monitor, high-availability
 - actions during probe 16
 - clear errors 27
 - defined 7
 - delete 27
 - disable 27
 - enable 27
 - event severity 21
 - probe configuration 17
 - probe timeout 17

- scripts 20
- view on Management Console 15
- mxnlmconfig command 37

N

- NFS clients
 - failover 7
 - filesystem access 6
 - timeout 36
- NLM locking protocol 37

O

- objects
 - Export Groups 3
 - export records 4
 - Virtual NFS Services 4

P

- probe, Export Group
 - failover behavior 18
 - frequency 17
 - timeout 17
- PSFS filesystems
 - mount options 8

R

- Recovery script 20

S

- scripts
 - custom 20
 - ordering of 22
- Start script 20
- status
 - Export Group 23
 - summary report 23
- Stop script 20

T

- technical support, HP v

V

- Virtual NFS Service
 - add 30
 - defined 6
 - delete 35
 - disable 34
 - enable 34
 - failover 7
 - guidelines for creating 29
 - object definition 4
 - policy, fallback 31
 - primary and backup interfaces 31
 - rehost 32